

Announcing the IA-64 Architecture

Hans Mulder

Lead Architect
Intel Corporation

Jerry Huck

Manager and Lead Architect
Hewlett Packard Co.

Introduction by: **Albert Yu**

Senior Vice President and General Manager
Microprocessor Products Group
Intel Corporation



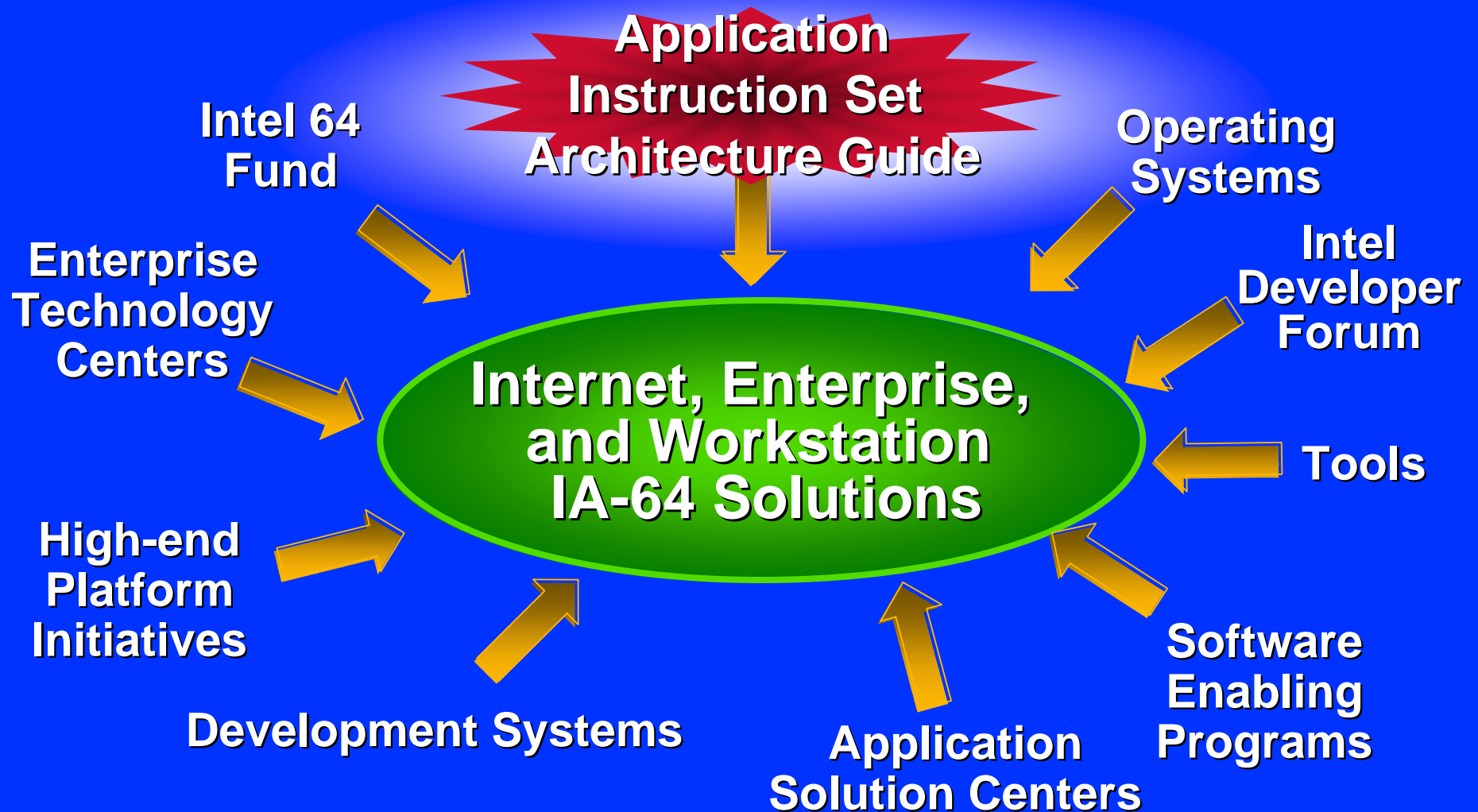
Agenda

- Introduction
- IA-64 Architecture Announcement
- IA-64 - Inside the Architecture
- Features for E-business
- Features for Technical Computing
- Summary

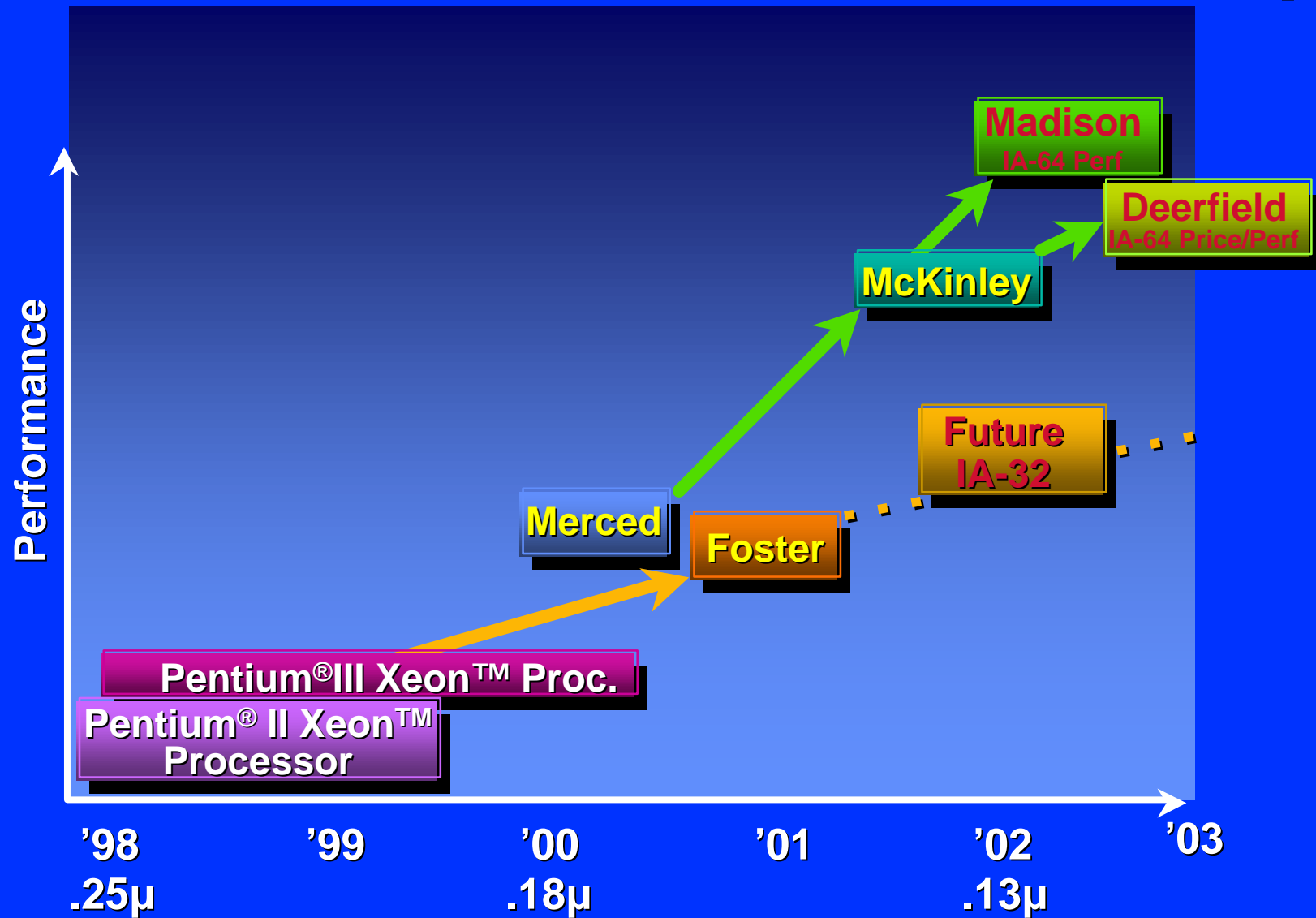
IA-64: A New Computing Era

- **Most significant architecture advancement since 32-bit computing with the 80386**
 - 80386: multi-tasking, advances from 16 bit to 32 bit
 - Merced: explicit parallelism, advances from 32 bit to 64 bit
- **Application Instruction Set Architecture Guide**
 - Complete disclosure of IA-64 application architecture
- **Result of the successful collaboration between Intel and HP**

Creating Complete IA-64 Solutions



IA Server/Workstation Roadmap

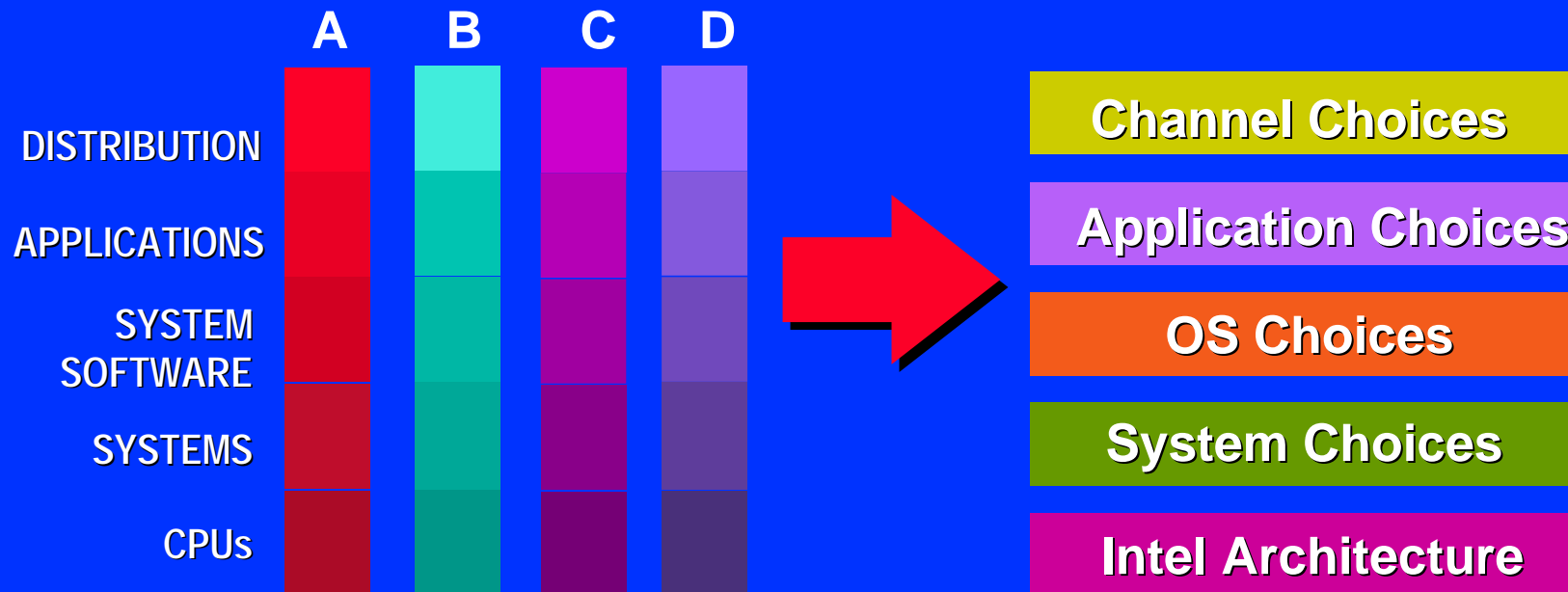


IA-64 starts with Merced processor

All dates specified are target dates provided for planning purposes only and are subject to change.

IA-64 Architecture Announcement

IA Changing the Face of High End Computing



“Vertical Market Structure”

- Limited Compatibility
- Few Choices
- Proprietary business

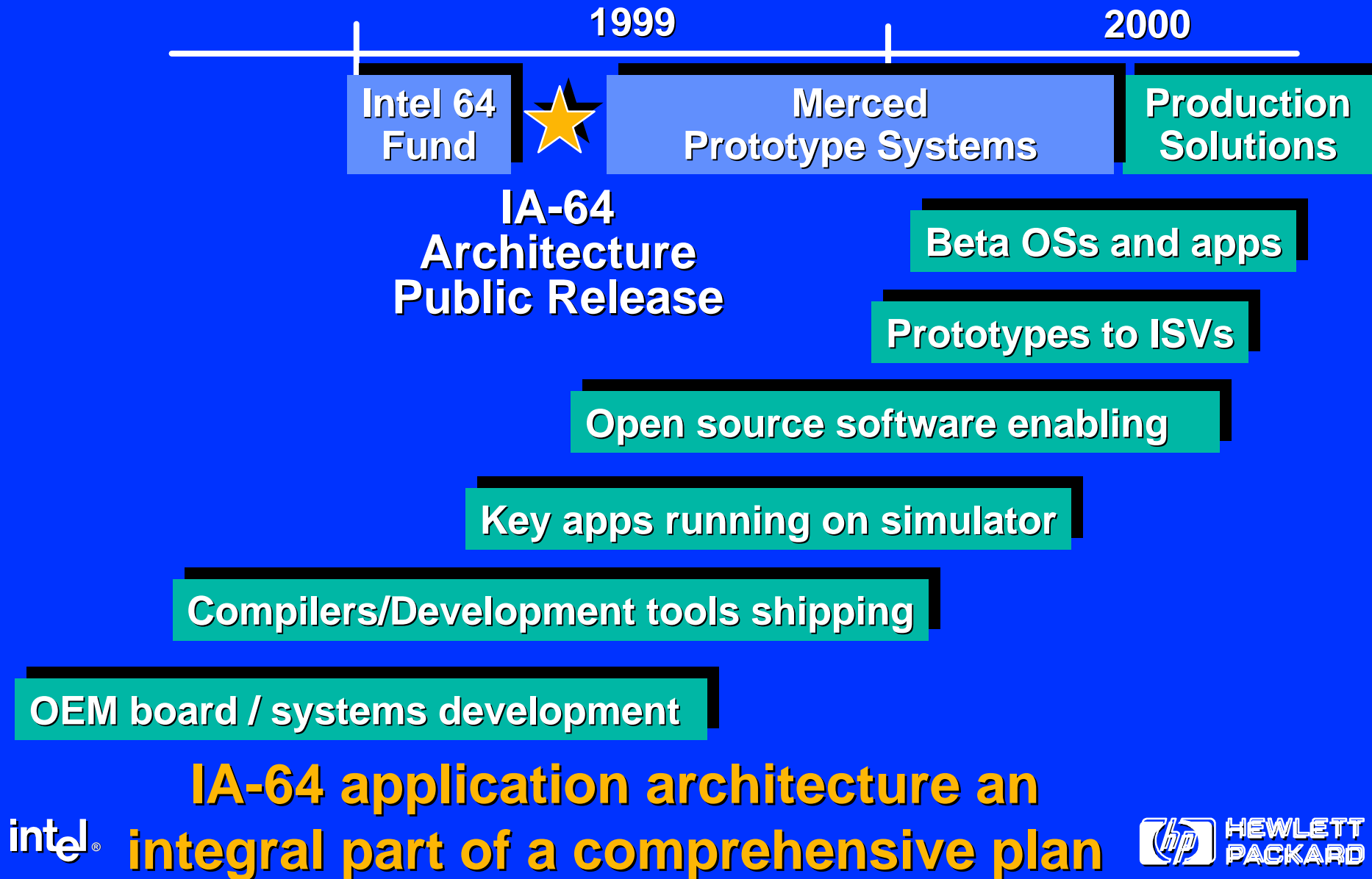
“Horizontal Market Structure”

- Highly Interoperable
- Many Choices
- Volume economics

**Unifying high end computing with a
common infrastructure**



Merced Industry Rollout



IA-64 Application Architecture

- **Application instructions and opcodes**
 - Instructions available to an application programmer
 - Machine code for these instructions
- **Unique architecture features & enhancements**
 - Explicit parallelism and templates
 - Predication, speculation, memory support, and others
 - Floating-point and multimedia architecture
- **IA-64 resources available to applications**
 - Large, application visible register set
 - Rotating registers, register stack, register stack engine
- **IA-32 & PA-RISC compatibility models**

Details now available to the broad industry

Today's Architecture Challenges

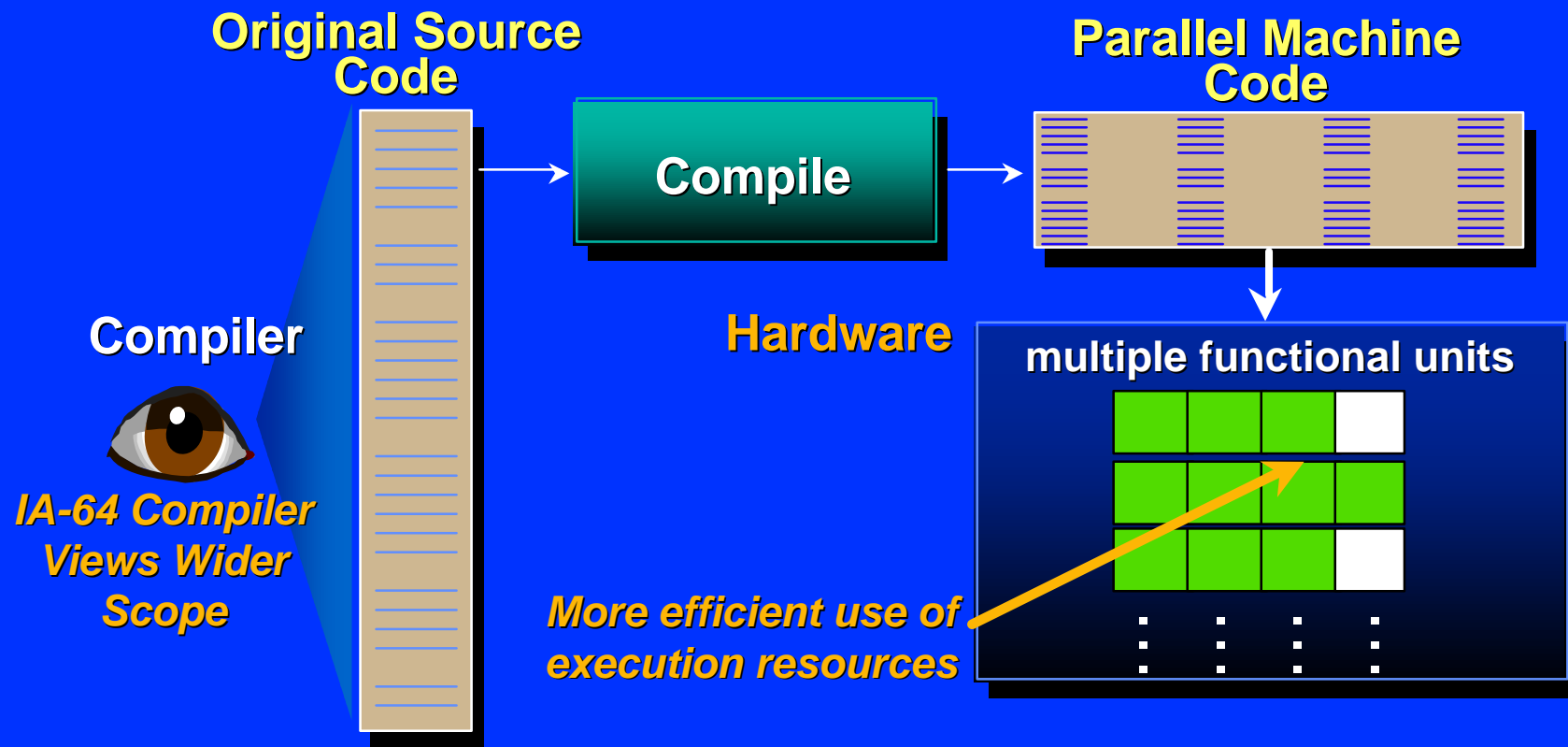
- Performance barriers :
 - Memory latency
 - Branches
 - Loop pipelining and call / return overhead
- Headroom constraints :
 - Hardware-based instruction scheduling
 - Unable to efficiently schedule parallel execution
 - Resource constrained
 - Too few registers
 - Unable to fully utilize multiple execution units
- Scalability limitations :
 - Memory addressing efficiency

IA-64 Mission

- Overcome the limitations of today's architectures
- Provide world-class floating-point performance
- Support large memory needs with 64-bit addressability
- Protect existing investments
 - Full binary compatibility with existing IA-32 instructions in hardware
 - Full binary compatibility with PA-RISC instructions through software translation
- Support growing high-end application workloads
 - E-business and internet applications
 - Scientific analysis and 3D graphics

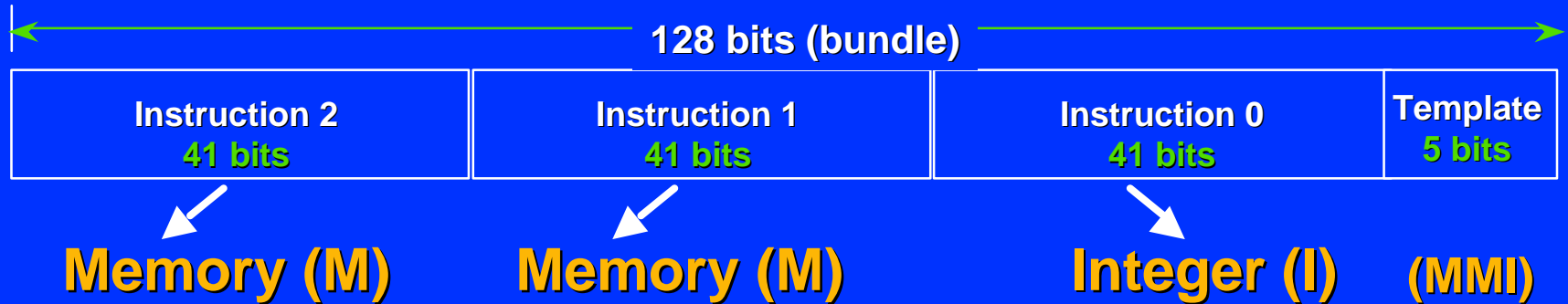
Define the next generation computer architecture

IA-64 Architecture : Explicit Parallelism



**Fundamental design philosophy
enables new levels of headroom**

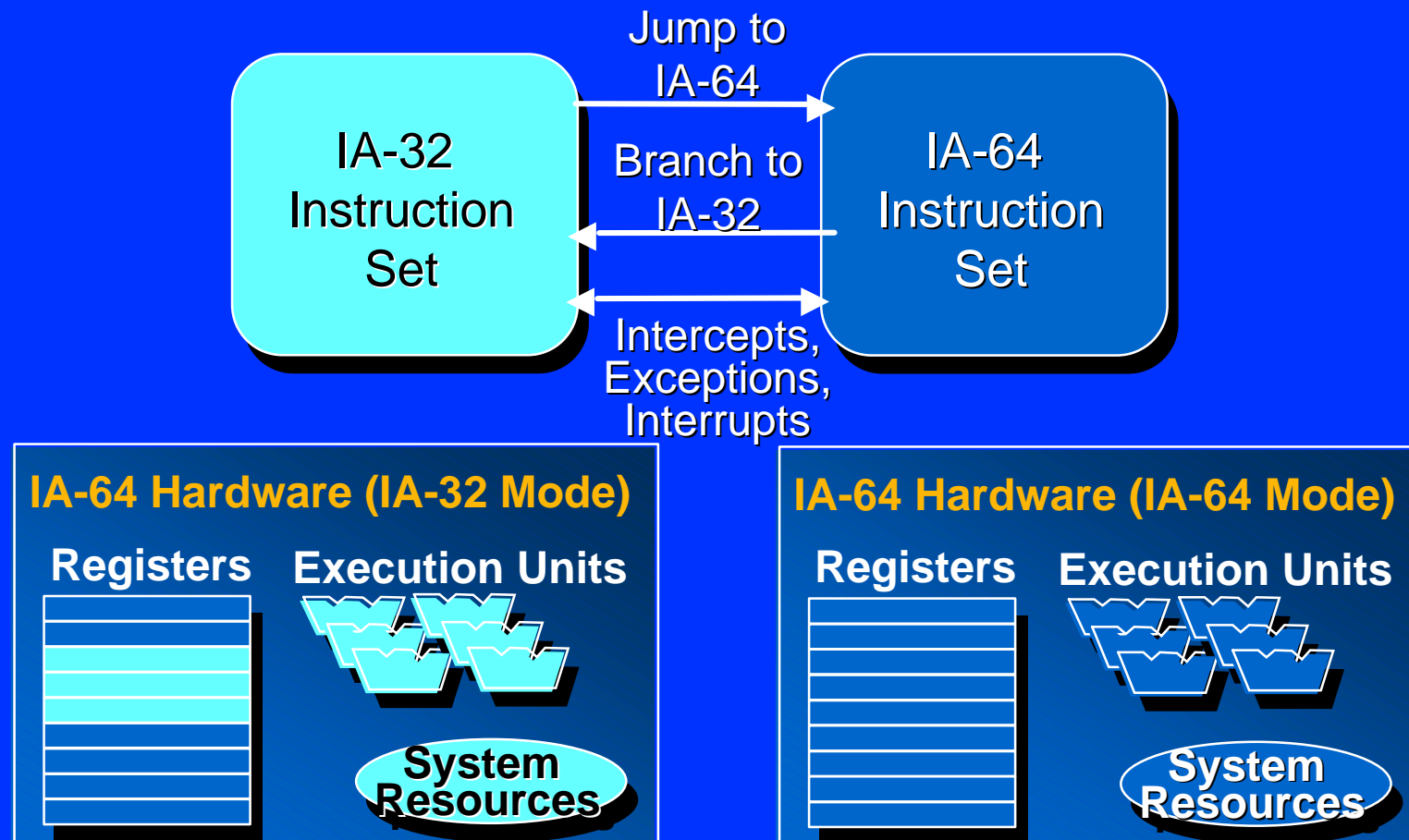
IA-64 : Explicitly Parallel Architecture



- IA-64 template specifies
 - The type of operation for each instruction
 - MFI, MMI, MII, MLI, MIB, MMF, MFB, MMB, MBB, BBB
 - Intra-bundle relationship
 - M / MI or MI / I
 - Inter-bundle relationship
- Most common combinations covered by templates
 - Headroom for additional templates
- Simplifies hardware requirements
- Scales compatibly to future generations

M=Memory
F=Floating-point
I=Integer
L=Long Immediate
B=Branch

Full Binary IA-32 Instruction Compatibility



- IA-32 instructions supported through shared hardware resources
- Performance similar to volume IA-32 processors

Preserves existing software investments

Full Binary Compatibility for PA-RISC

- **Transparency:**

- Dynamic object code translator in HP-UX automatically converts PA-RISC code to native IA-64 code
- Translated code is preserved for later reuse

- **Correctness:**

- Has passed the same tests as the PA-8500

- **Performance:**

- Close PA-RISC to IA-64 instruction mapping
- Translation on average takes 1-2% of the time Native instruction execution takes 98-99%
- Optimization done for wide instructions, predication, speculation, large register sets, etc.
- PA-RISC optimizations carry over to IA-64

High Performance Computing Applications

E-business servers

- Large number of users
- Large databases
- High availability
- Secure environment



Workstations and high performance technical computing

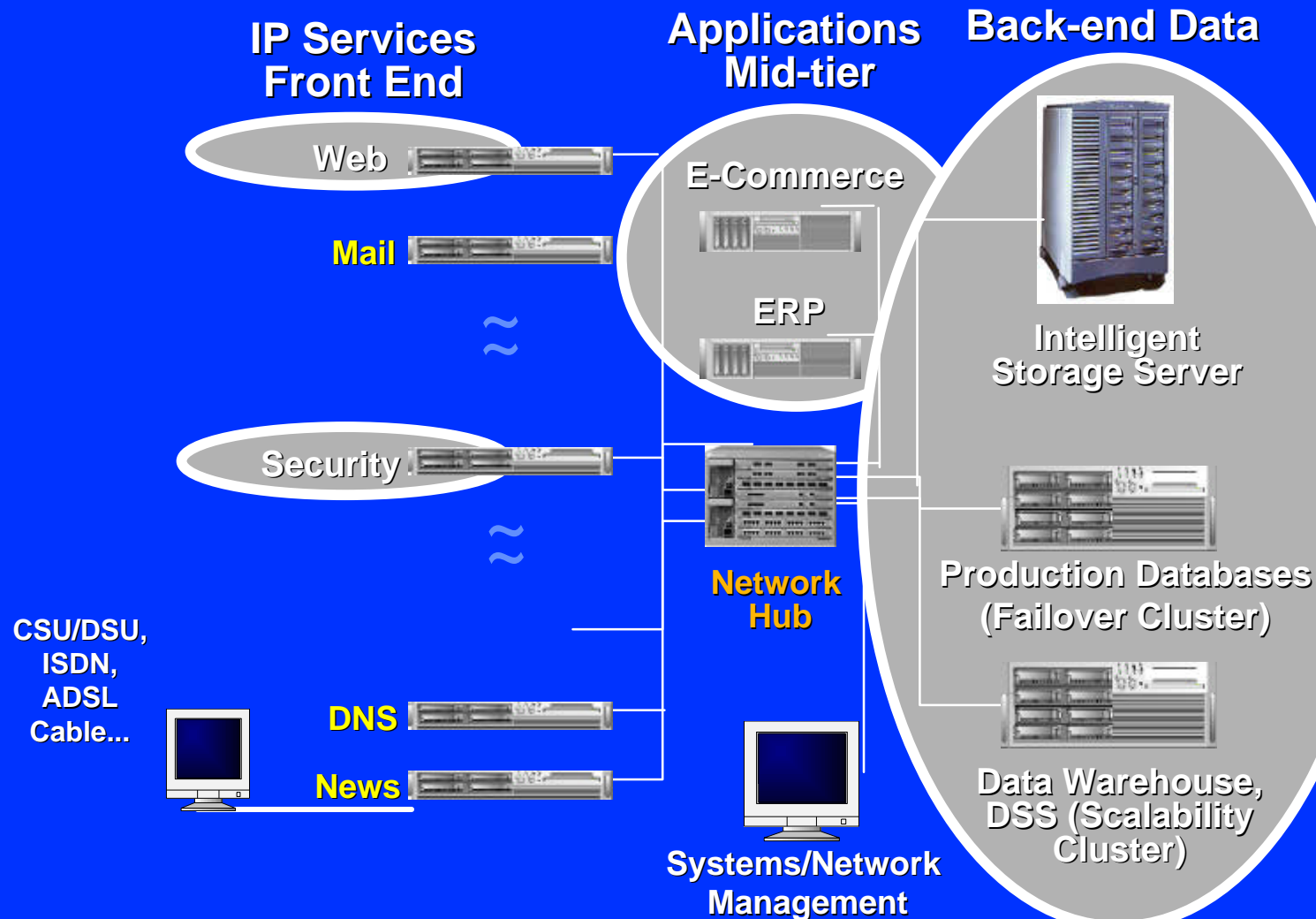
- Digital content creation
- Design engineering (EDA, MDA, etc)
- Scientific / financial analysis

**IA-64 architecture optimized for these
high growth applications**



E-Business Environment

IA-64 focus area



E-business is compute- intensive requiring security and support for large databases



IA-64 for High Performance Databases

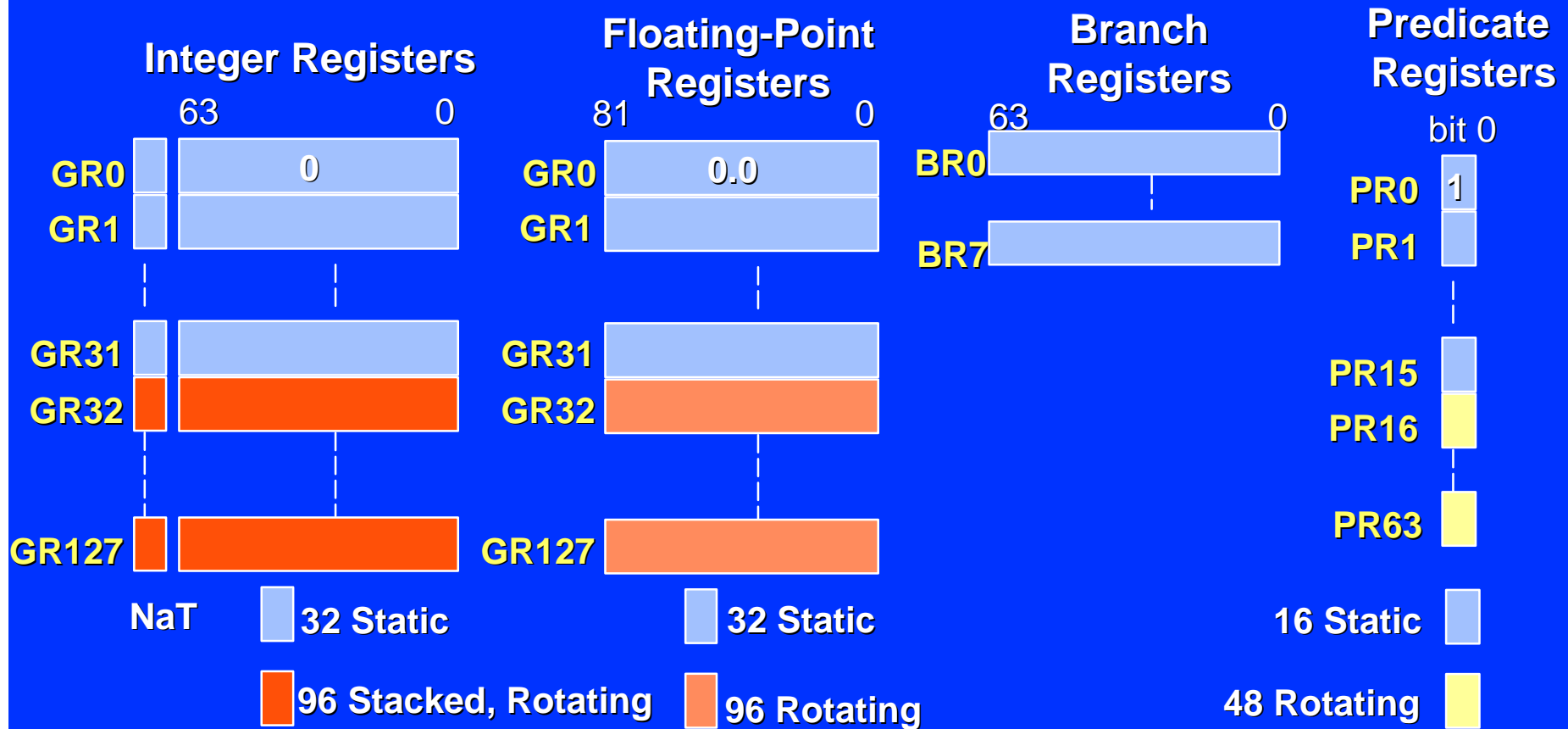
- **Number of branches in large server apps overwhelm traditional processors**
 - IA-64 predication removes branches, avoids mispredicts
- **Environments with a large number of users require high performance**
 - IA-64 uses speculation to reduce impact of memory latency
 - Significant benefit to large databases with many cache accesses
 - 64-bit addressing enables systems with very large virtual and physical memory

Middle Tier Application Needs

- Mid-tier applications (ERP, etc.) have diverse code requirements
 - Integer code with many small loops
 - Significant call / return requirements (C++, Java)
- IA-64's unique register model supports these various requirements
 - Large register file provides significant resources for optimized performance
 - Rotating registers enables efficient loop execution
 - Register stack to handle call-intensive code

IA-64 resources enable optimization for a variety of application requirements

IA-64's Large Register File

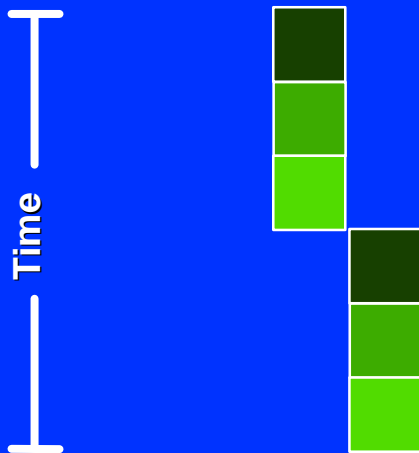


Large number of registers enables flexibility and performance

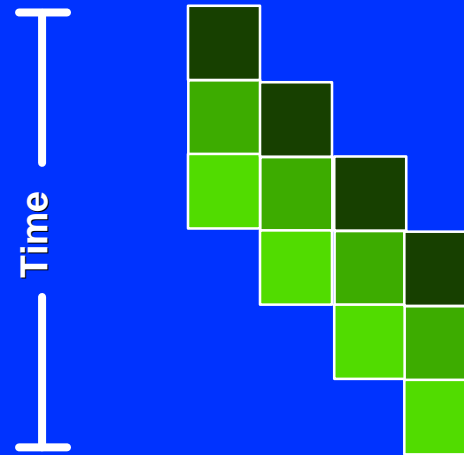
Software Pipelining via Rotating Registers

- Software pipelining - improves performance by overlapping execution of different software loops - execute more loops in the same amount of time

Sequential Loop Execution



Software Pipelining Loop Execution

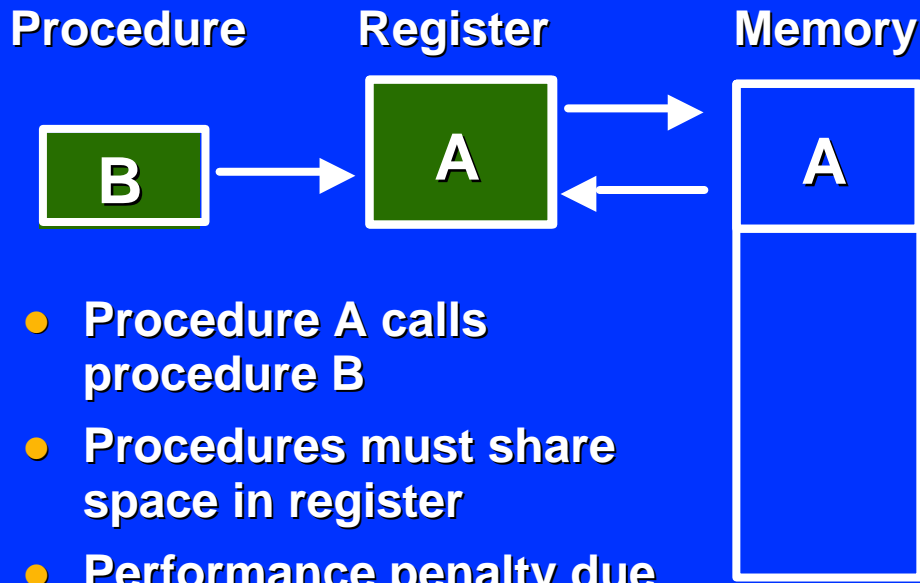


- Traditional architectures need complex software loop unrolling for pipelining
 - Results in code expansion --> Increases cache misses --> Reduces performance
- IA-64 utilizes rotating registers to achieve software pipelining
 - Avoids code expansion --> Reduces cache misses --> Higher performance

**IA-64 rotating registers enable
optimized loop execution**

Traditional Register Models

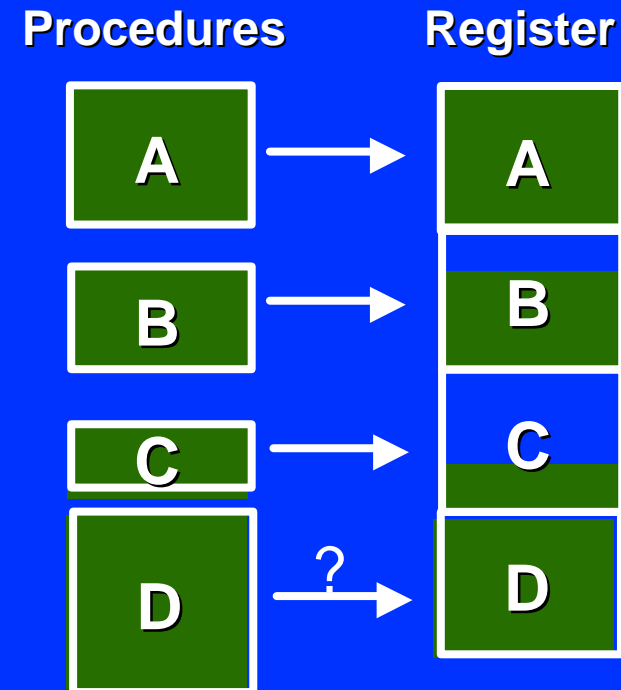
Traditional Register Models



- Procedure A calls procedure B
- Procedures must share space in register
- Performance penalty due to register save / restore

IA-64 significantly improves upon this

Traditional Register Stacks



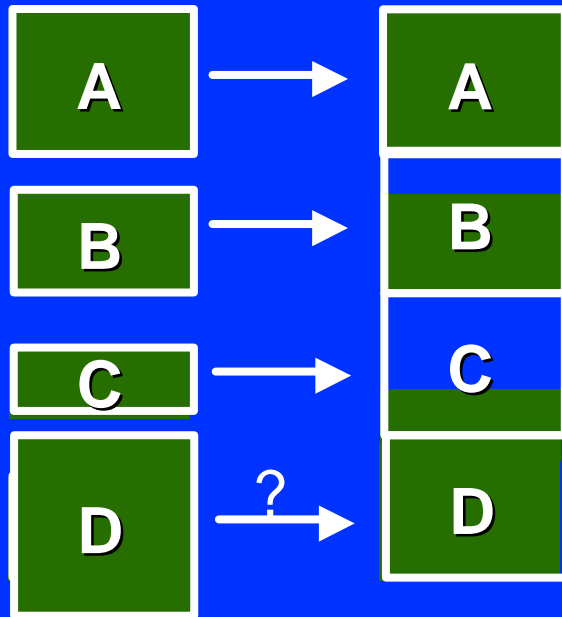
- Eliminate the need for save / restore by reserving fixed blocks in register
- However, fixed blocks waste resources

IA-64 Register Stack

Traditional Register Stacks

Procedures

Register

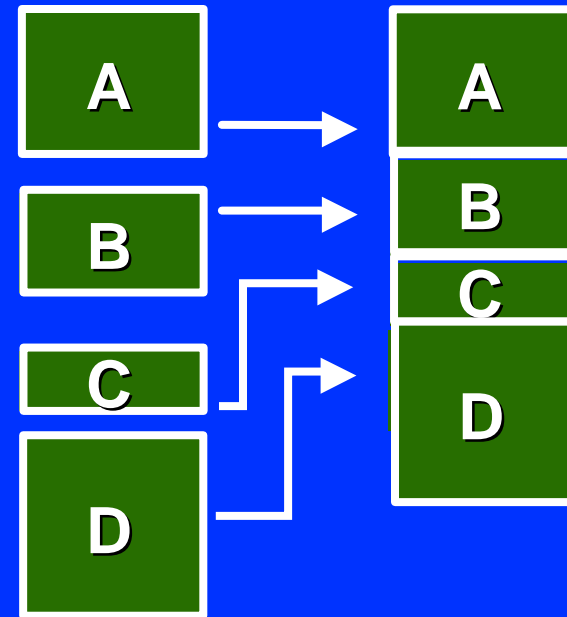


- Eliminate the need for save / restore by reserving fixed blocks in register
- However, fixed blocks waste resources

IA-64 Register Stack

Procedures

Register

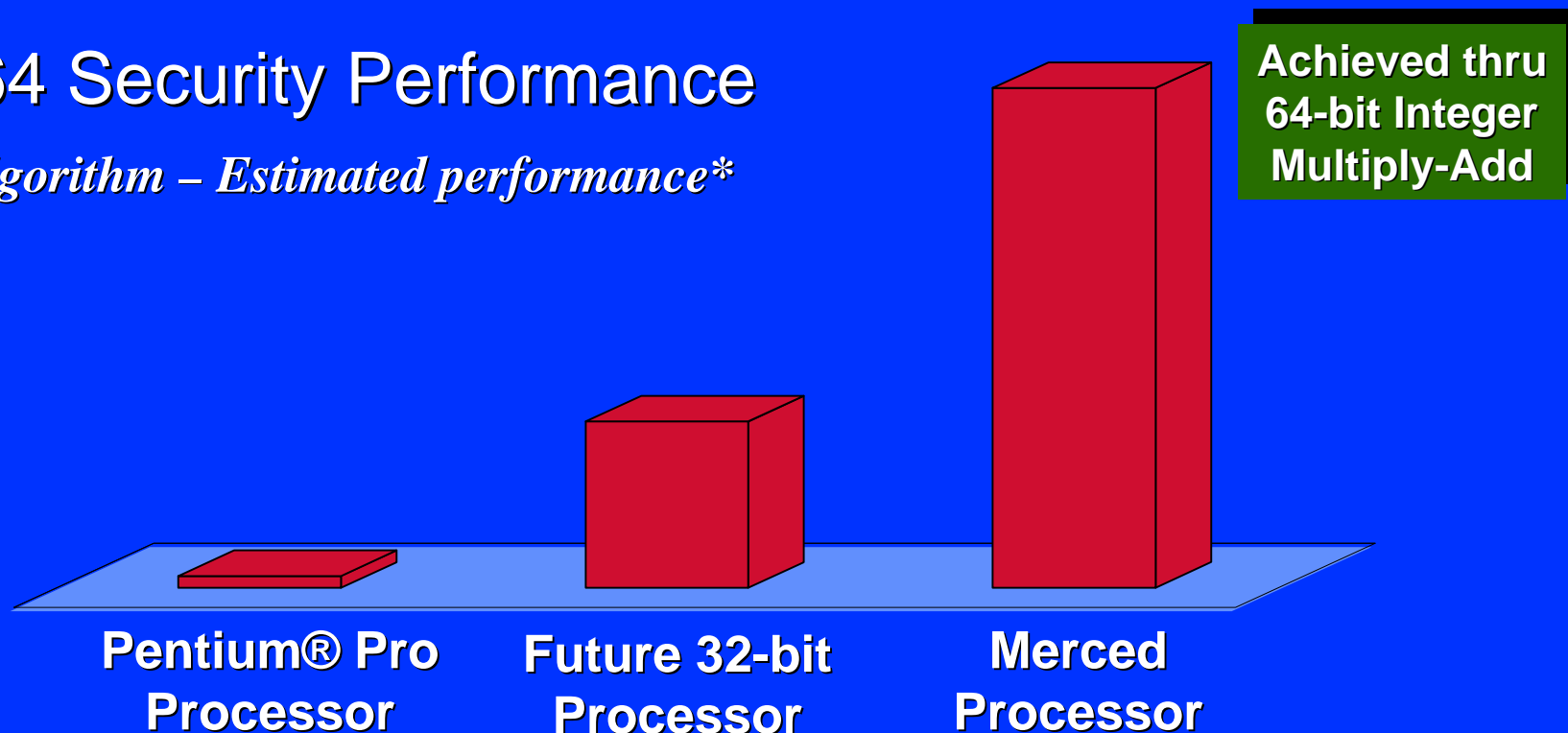


- IA-64 able to reserve variable block sizes
- No wasted resources

**IA-64 combines high performance
and high efficiency**

IA-64 Security Performance for E-Business

IA-64 Security Performance
*RSA Algorithm – Estimated performance**



IA-64 delivers secure transactions to more users



**Intel estimates*



* All third party marks, brands, and names are the property of their respective owners 24

Delivery of Streaming Media

- Audio and video functions regularly perform the same operation on arrays of data values
 - IA-64 manages its resources to execute these functions efficiently
 - Able to manage general register's as 8x8, 4x16, or 2x32 bit elements
 - Multimedia operands/results reside in general registers
- IA-64 accelerates compression / decompression algorithms
 - Parallel ALU, Multiply, Shifts
 - Pack/Unpack; converts between different element sizes.
- Fully compatible with IA-32 MMX technology, Streaming SIMD Extensions and PA-RISC MAX2

IA-64 resources and parallelism enables efficient delivery of rich web content

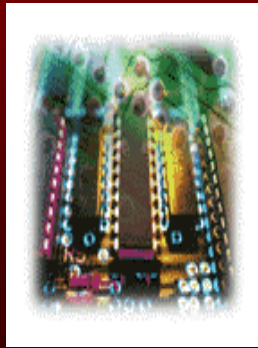
Technical Computing Environment

- Rendering
- Editing
- 3D Animation



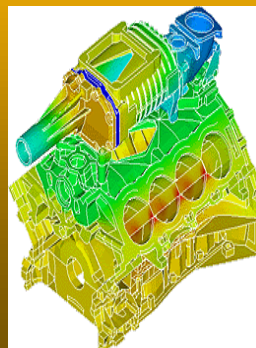
DCC

- Verification
- Synthesis
- DRC



EDA

- FEA
- Modeling
- Hi-end CAE



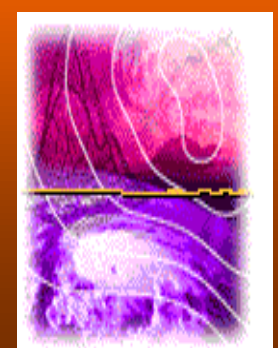
MDA

- Equity
- Treasury
- Risk Analysis



Finance

- CFD
- GIS
- Molecular



**Scientific
Analysis**

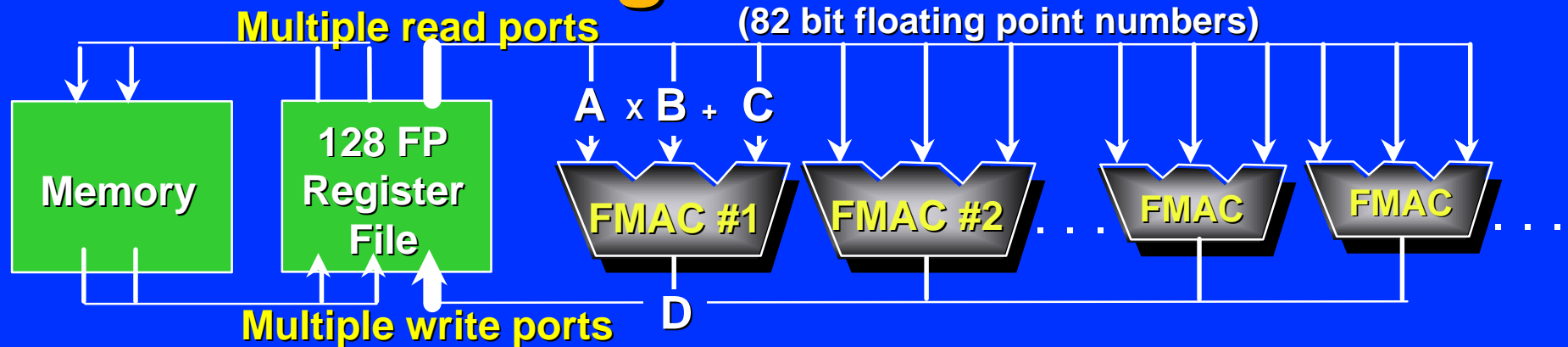
**High performance
floating-point is key**

IA-64 for Scientific Analysis

- **Variety of software optimizations supported**
 - Load double pair : doubles bandwidth between L1 & registers
 - Full predication and speculation support
 - NaT Value to propagate deferred exceptions
 - Alternate IEEE flag sets allow preserving architectural flags
 - Software pipelining for large loop calculations
- **High precision & range internal format : 82 bits**
 - Mixed operations supported: single, double, extended, and 82-bit
 - Interfaces easily with memory formats
 - Simple promotion/demotion on loads/stores
 - Iterative calculations converge faster
 - Ability to handle numbers much larger than RISC competition without overflow

High performance & High precision

IA-64 Floating-Point Architecture



- **128 registers**
 - Allows parallel execution of multiple floating-point operations
- **Simultaneous Multiply - Accumulate (FMAC)**
 - 3-input, 1-output operation : $a * b + c = d$
 - Shorter latency than independent multiply and add
 - Greater internal precision and single rounding error

**Resourced for scientific
analysis and 3D graphics**

IA-64 3D Graphics Capabilities

- Many geometric calculations (transforms and lighting) use 32-bit floating-point numbers
- IA-64 configures registers for maximum 32-bit floating-point performance
 - Floating-point registers treated as 2x32 bit single precision registers
 - Able to execute fast divide
 - Achieves up to 2X performance boost in 32-bit data floating-point operations
- Full support for Pentium® III processor Streaming SIMD Extensions (SSE)

IA-64 enables world-class GFLOPs performance

Memory Support for High Performance Technical Computing

- Scientific analysis, 3D graphics and other technical workloads tend to be predictable & memory bound
- IA-64 data pre-fetching of operations allows for fast access of critical information
 - Reduces memory latency impact
- IA-64 able to specify cache allocation
 - Cache hints from load / store operations allow data to be placed at specific cache level
 - Efficient use of caches, efficient use of bandwidth

Reduces the memory bottleneck

IA-64 Features

Explicit Parallelism : compiler / hardware synergy

Register Model : large register file, rotating registers, register stack engine

Floating Point Architecture : extended precision calculations, 128 registers, FMAC, SIMD

Multimedia Architecture : parallel arithmetic, parallel shift, data arrangement instructions

Memory Management : 64-bit addressing, speculation, memory hierarchy control

Compatibility : full binary compatibility with existing IA-32 instructions in hardware, PA-RISC through software translation

Function

Executes more instructions in the same amount of time

Able to optimize for scalar and object oriented applications

High performance 3D graphics and scientific analysis

Improves calculation throughput for multimedia data

Manages large amounts of memory, efficiently organizes data from / to memory

Existing software runs seamlessly

Benefits

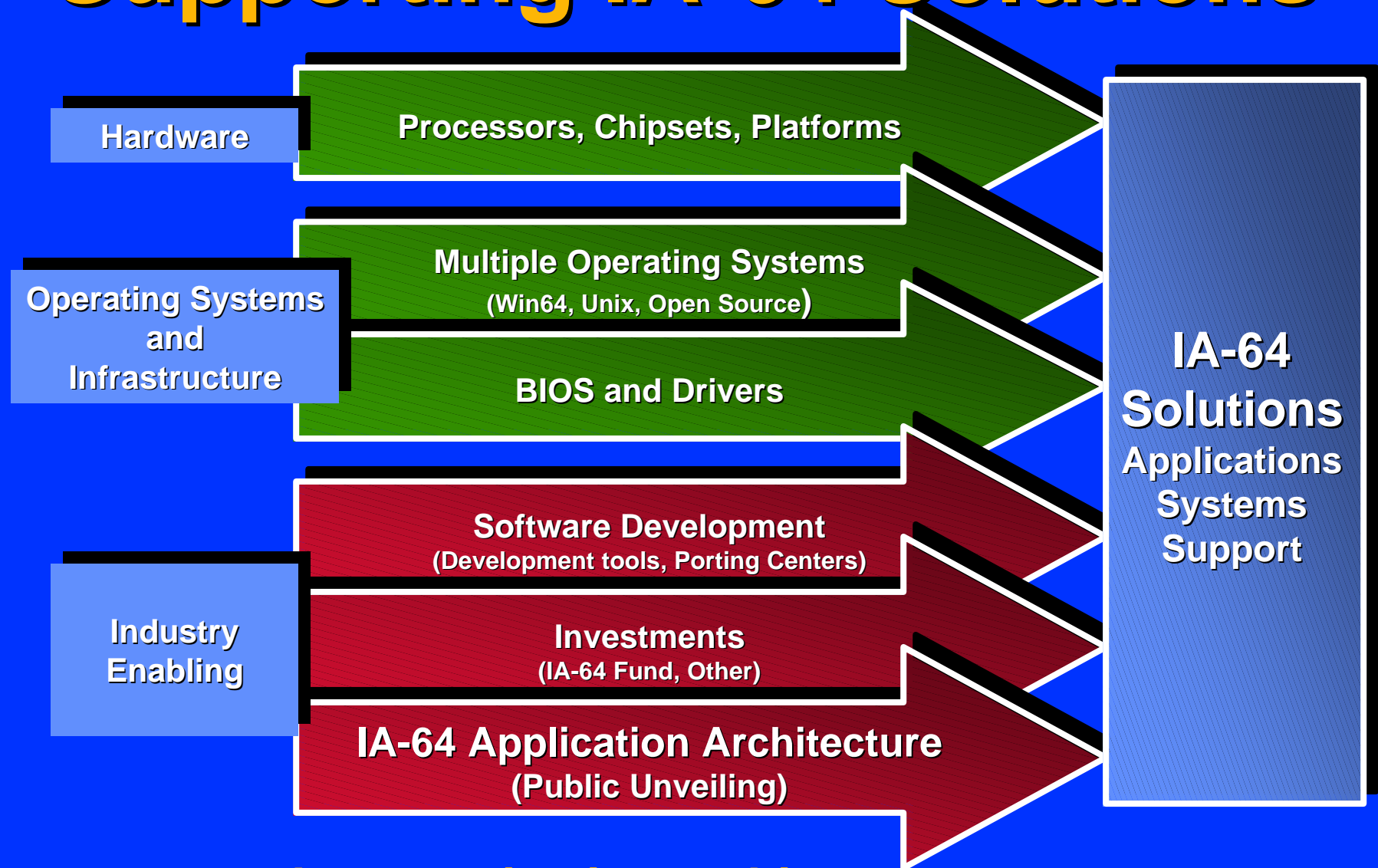
- Maximizes headroom for the future
- World-class performance for complex applications
- Enables more complex scientific analysis
- Faster digital content creation and rendering
- Efficient delivery of rich Web content
- Increased architecture & system scalability
- Preserves investment in existing software

IA-64 Details Made Public

- **IA-64 Application ISA Guide (AIG)**
 - Application instructions and machine code
 - Application programming model
 - Unique architecture features & enhancements
- **Provides understanding of IA-64 for the broad industry**
 - Features and benefits for key applications
 - Insight into techniques for optimizing IA-64 solutions
- **IA-64 AIG and other developer information available 5/26**
 - <http://developer.intel.com/design/ia64/index.htm>
 - <http://www.hp.com/go/ia64>

Continuing to fuel IA-64 developer momentum

Supporting IA-64 Solutions



IA-64 application architecture an integral part of a comprehensive plan



Summary

- IA-64 represents the most significant architecture development since 80386
- IA-64 advances beyond the capabilities of traditional architectures
 - Compiler / hardware synergy, massive resources, headroom
- IA-64 provides features to benefit the high-end applications of the future
 - E-business
 - Technical computing
- Today's architecture unveiling is an additional element of the comprehensive IA-64 industry program

IA-64 begins with Merced